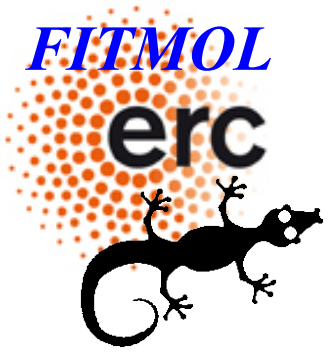


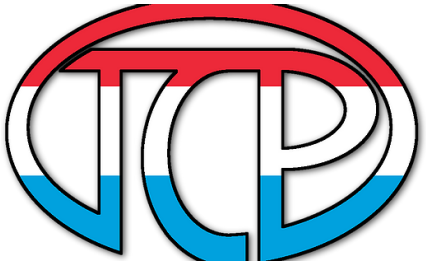
Navigating Chemical Compound Space … Directly and Inversely

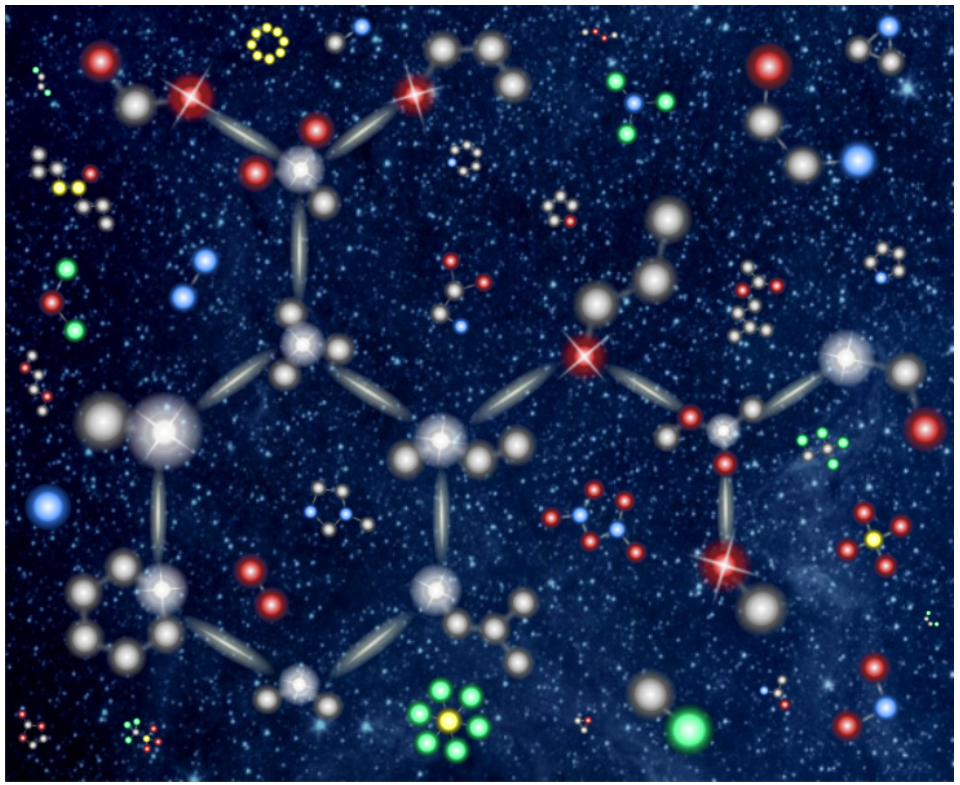
Alexandre Tkatchenko *et al.*

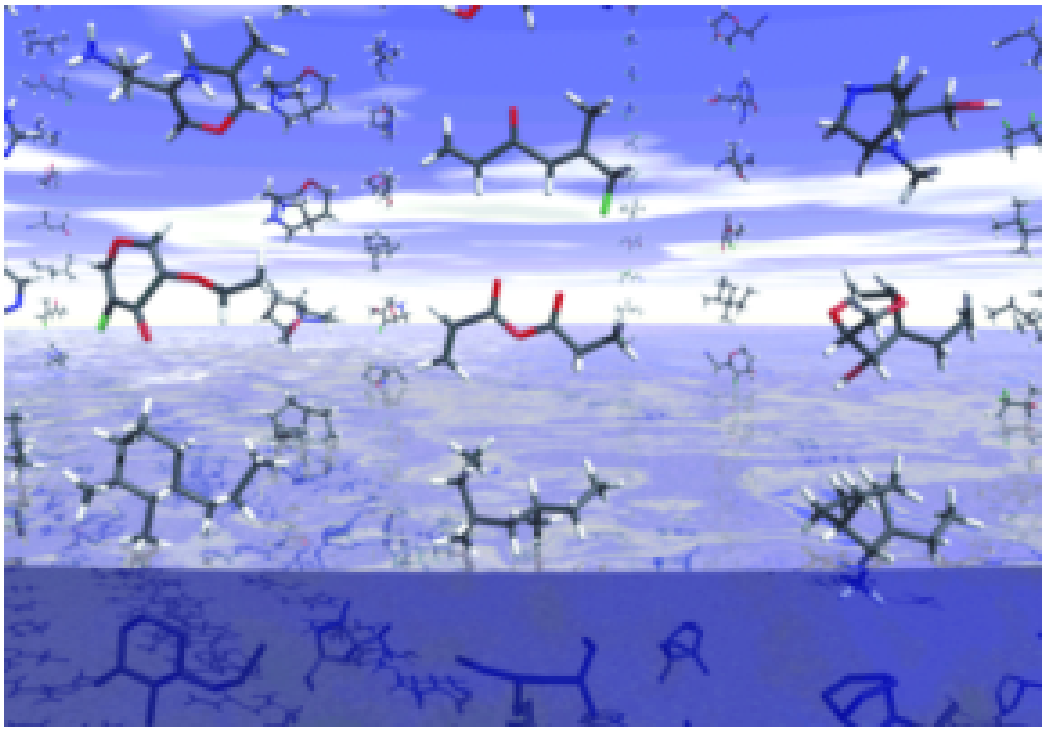
Chair for Theoretical Chemical Physics (TCP),
Dept of Physics and Materials Science (DPhyMS), *University of Luxembourg*

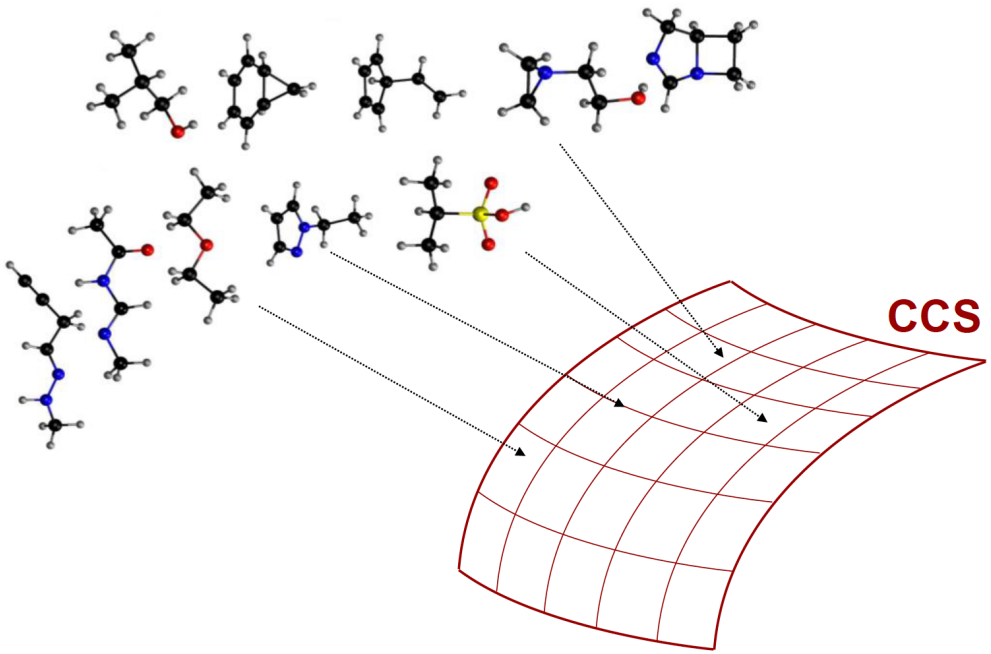
www.tcpunilu.com
alexandre.tkatchenko@uni.lu

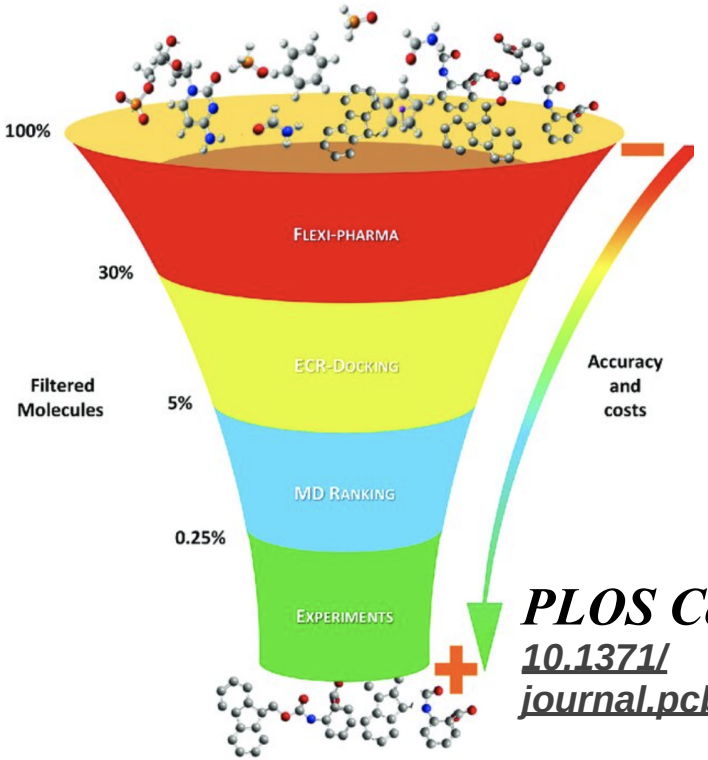
AIDD, Berlin





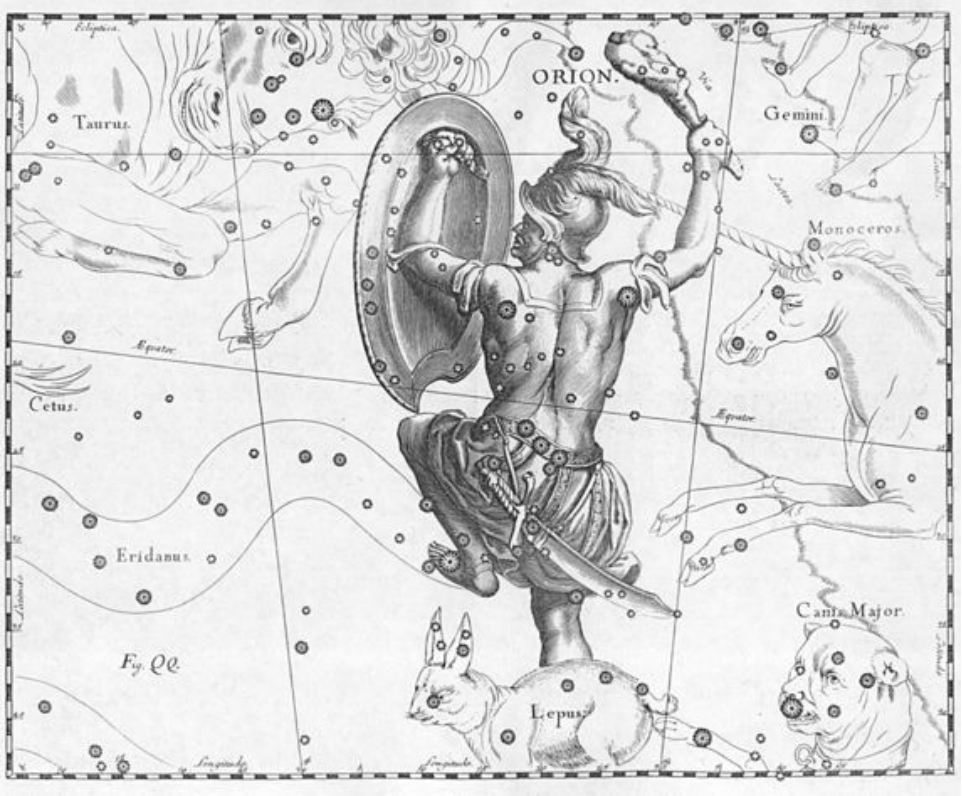




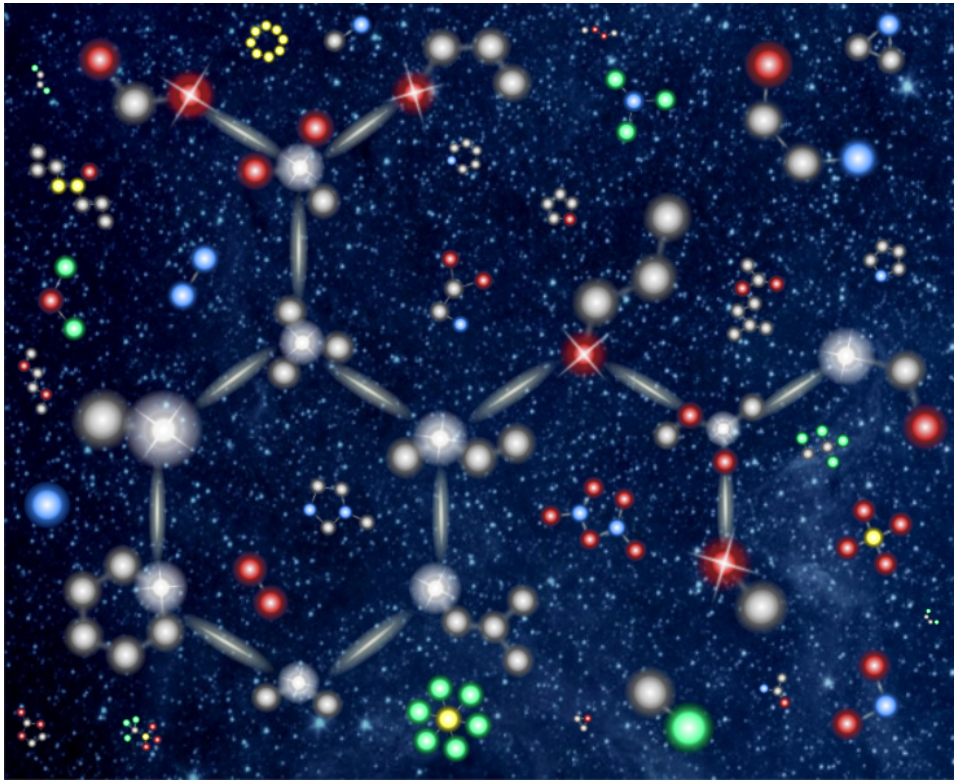


PLOS Comp. Bio.
10.1371/
journal.pcbi.1007898

Our Wildest Dream: Discovering "Constellations" in Chemical Space



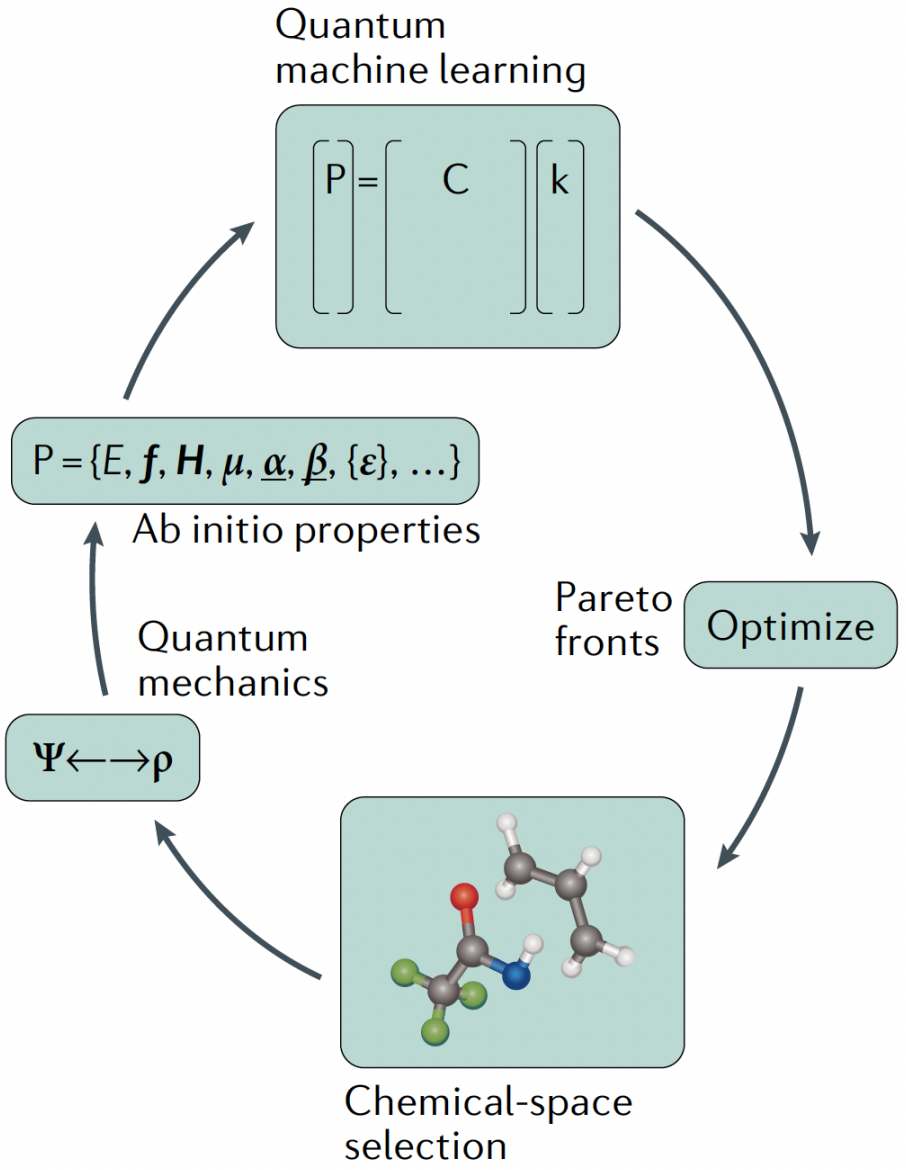
Constellation Orion, Credit: Wikipedia



"Constellation Aspirin"

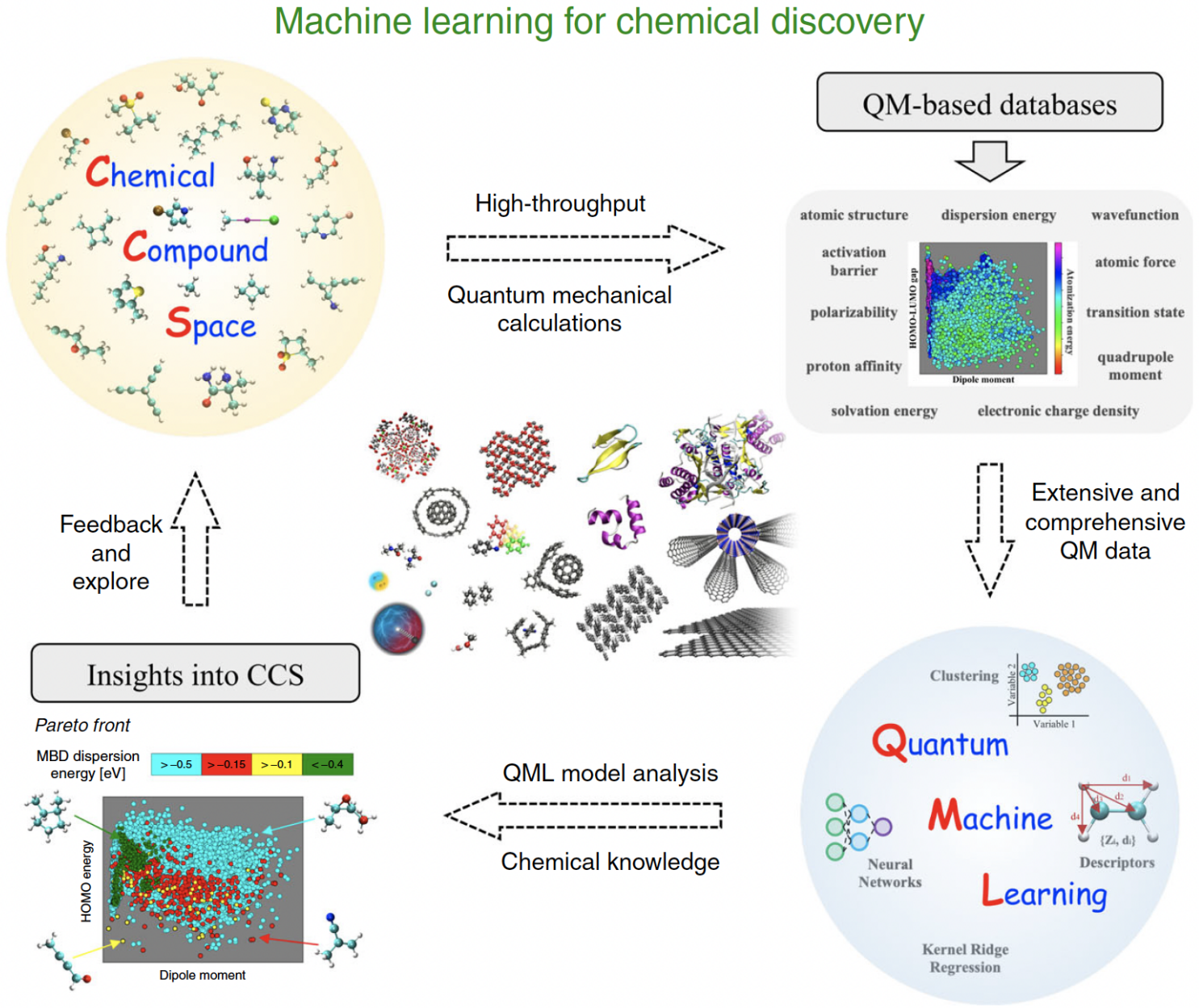
O. A. von Lilienfeld, K.-R. Mueller, and A. Tkatchenko, ***Nature Rev. Chem.*** 4, 347 (2020).

Chemical Compound Space: Designing Molecules



O. A. von Lilienfeld, K.-R. Mueller, and A. Tkatchenko,
Nature Rev. Chem. 4, 347 (2020).

Chemical Compound Space: Designing Molecules



A. Tkatchenko, ***Nature Commun.*** 11, 4125 (2020).

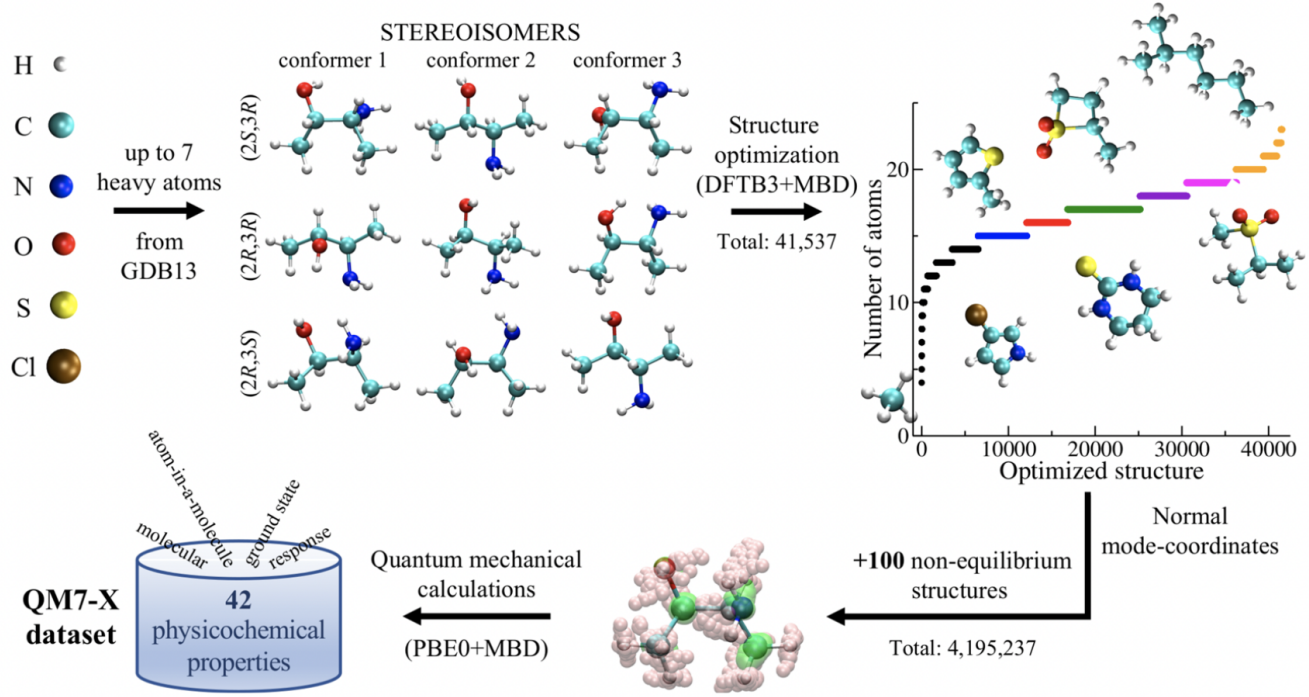
CCS of Small Molecules: QM7-X Dataset

Data Descriptor | Open access | Published: 02 February 2021

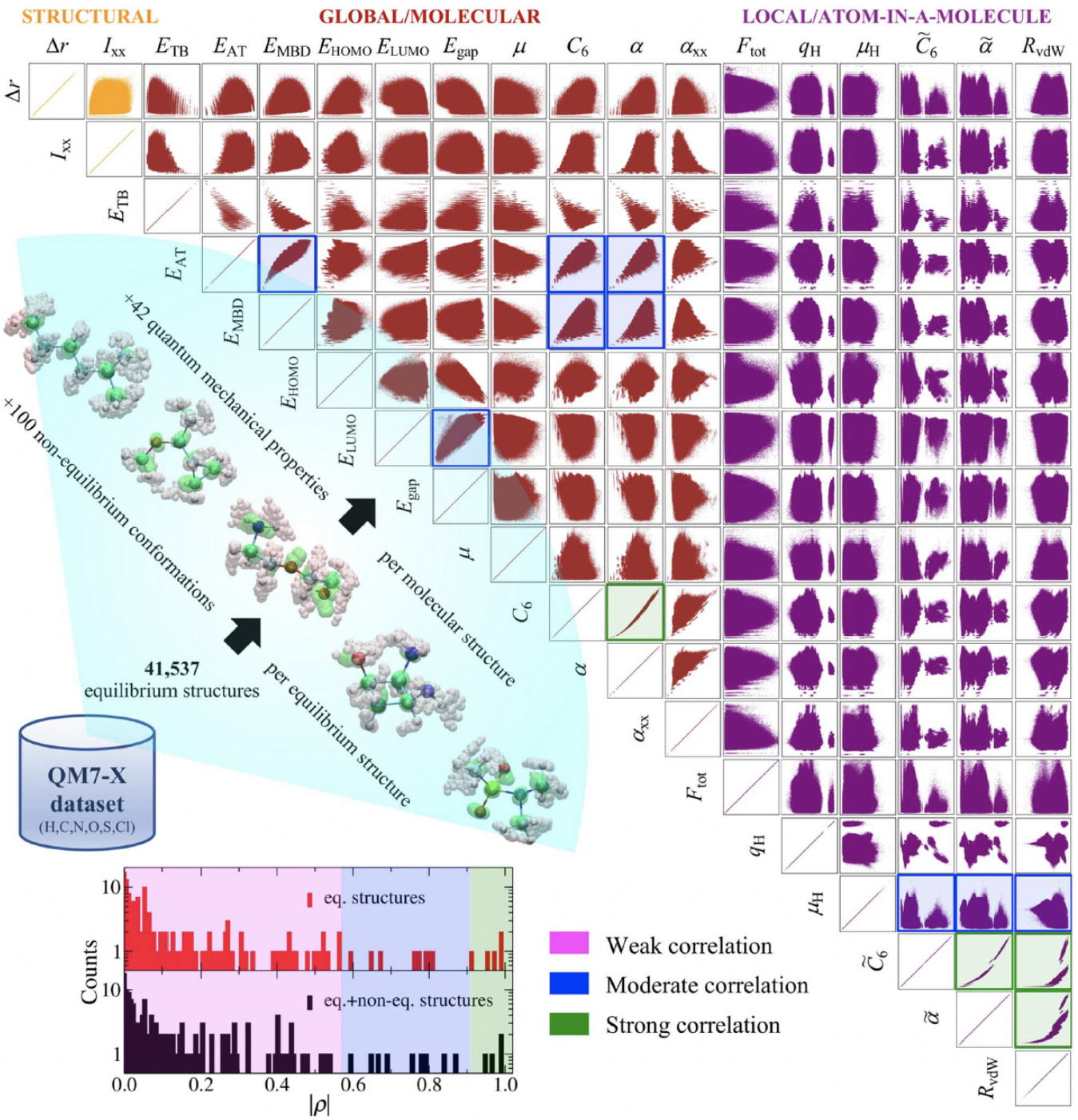
QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules

Johannes Hoja, Leonardo Medrano Sandonas, Brian G. Ernst, Alvaro Vazquez-Mayagoitia, Robert A. DiStasio Jr. ✉ & Alexandre Tkatchenko ✉

Scientific Data **8**, Article number: 43 (2021) | Cite this article



Chemical Compound Space: Freedom of Design



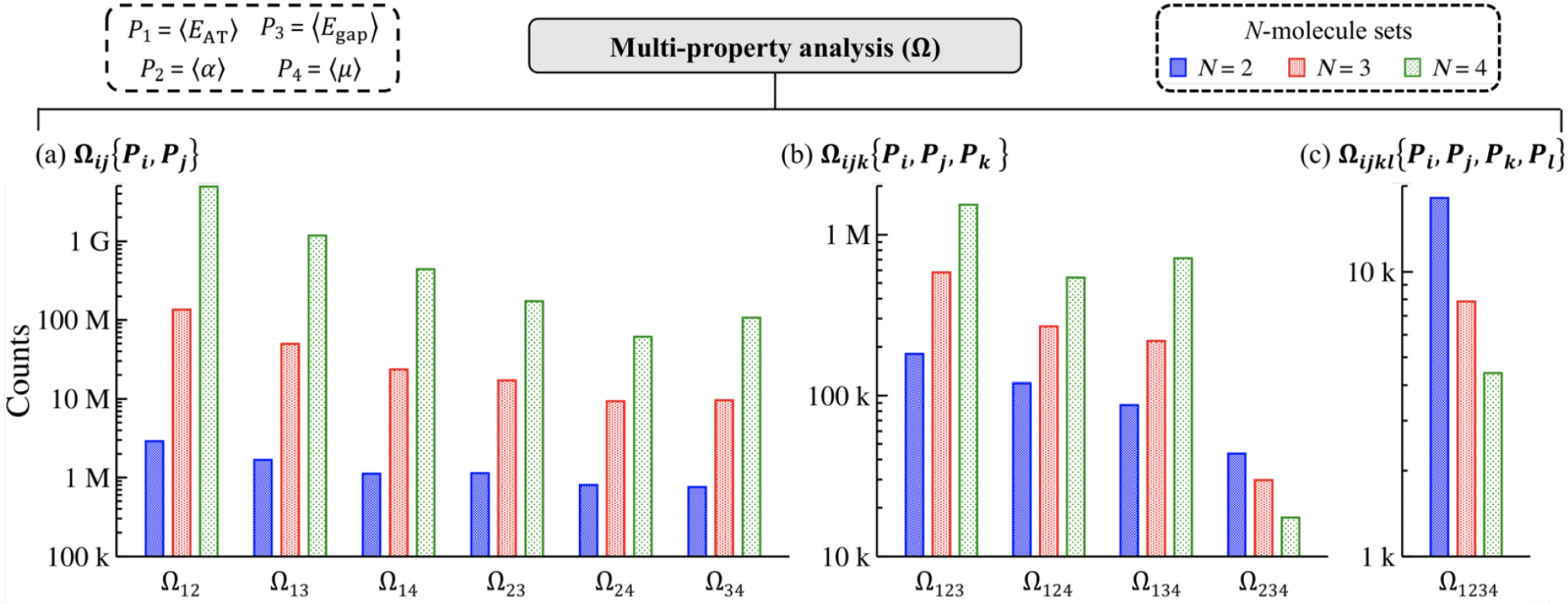
Chemical Compound Space: Freedom of Design

Cite this: DOI: 10.1039/d3sc03598k

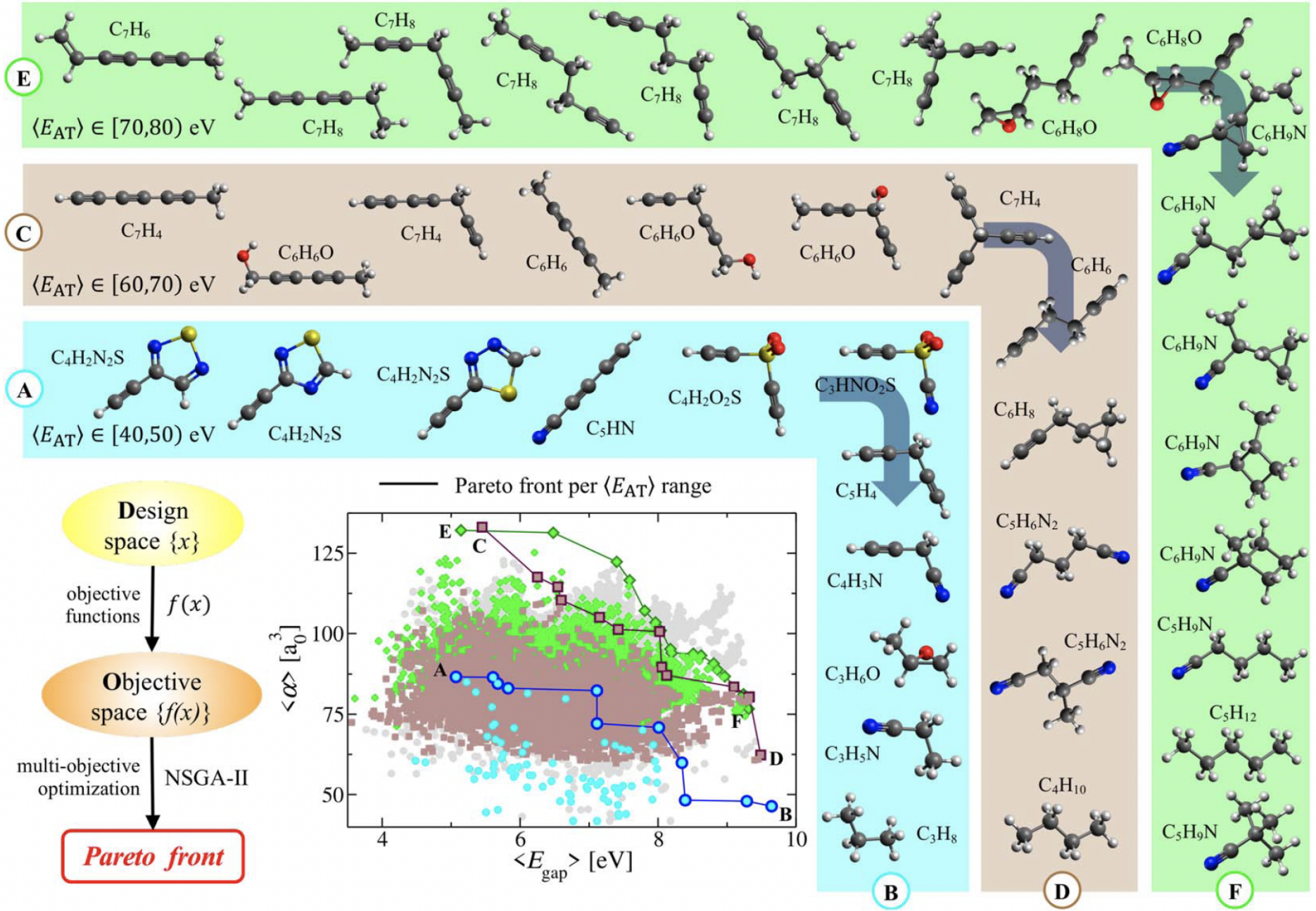
All publication charges for this article have been paid for by the Royal Society of Chemistry

"Freedom of design" in chemical compound space: towards rational *in silico* design of molecules with targeted quantum-mechanical properties†

Leonardo Medrano Sandonas, *[a] Johannes Hoja, [ab] Brian G. Ernst, [c] Álvaro Vázquez-Mayagoitia, [d] Robert A. DiStasio, Jr *[c] and Alexandre Tkatchenko *[a]



Chemical Compound Space: Pareto Fronts



L. Medrano Sandonas et al., ***Chem. Sci.*** 14, 10702 (2023).

(Baby)-Steps Towards Inverse Molecular Design

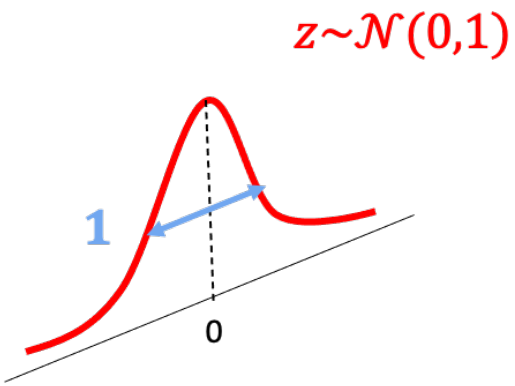
Can we parameterize the chemical space of structures with a select set of QM properties?

Goal:
ML model to map QM properties to structures

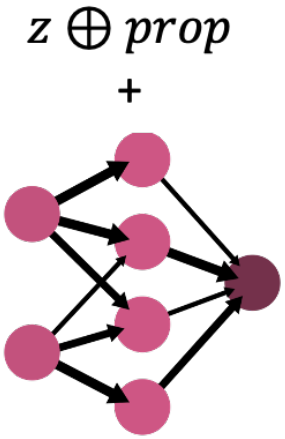
Towards Inverse Molecular Design

A closely related problem is already addressed by a plethora of conditional generative models following this scheme:

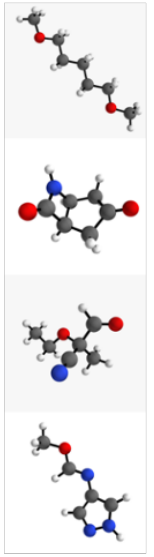
1) Sample a Normal distribution…



2)…condition on one/two properties…



3)…obtain some molecules



Gebauer et al. Nat Commun 2022 13, 973

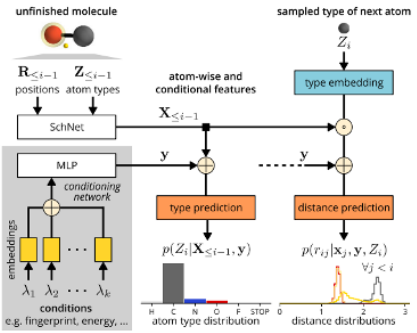
Towards Inverse Molecular Design

Two very successful examples with 3D structures…

https://doi.org/10.1038/s41467-022-28526-y OPEN

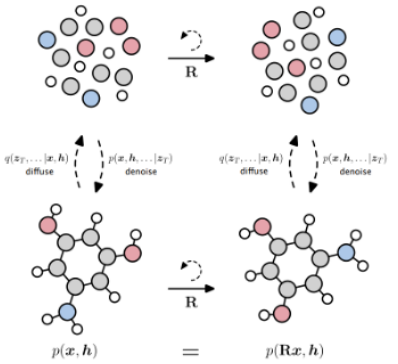
Inverse design of 3d molecular structures with conditional generative neural networks

Niklas W. A. Gebauer[1,2,3✉], Michael Gastegger[1,3], Stefaan S. P. Hessmann[1,2], Klaus-Robert Müller[1,2,4,5] & Kristof T. Schütt[1,2✉]



Equivariant Diffusion for Molecule Generation in 3D

Emiel Hoogeboom [*1] Victor Garcia Satorras [*1] Clément Vignac [*2] Max Welling [1]



…but they are not direct deterministic properties-structure mappings

Hogeboom et al. ARXIV.2203.17003 2022

Gebauer et al. Nat Commun 2022 13, 973

(Baby)-Steps Towards Inverse Molecular Design

Inverse Mapping of Quantum Properties to Structures for Chemical Space of Small Organic Molecules

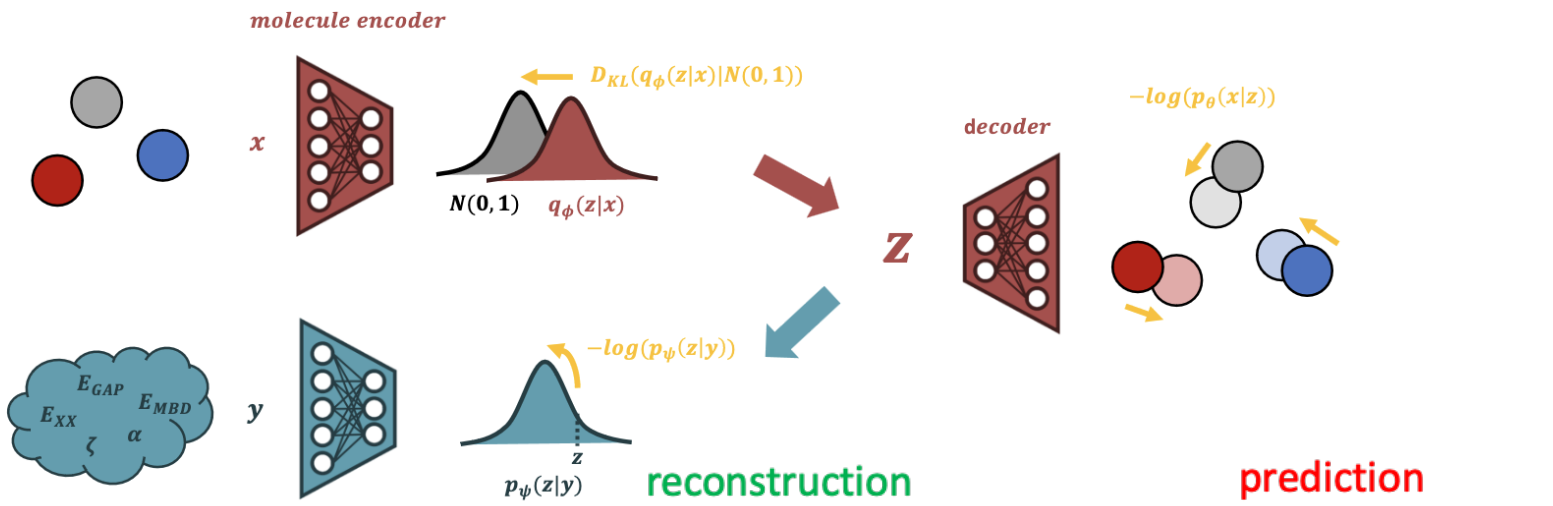
Alessio Fallani[1], Leonardo Medrano Sandonas[1,†], and Alexandre Tkatchenko[1]

[1]Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg.
[*]Corresponding authors: Alessio Fallani (alessio.fallani@uni.lu), Leonardo Medrano Sandonas (leonardo.medrano@tu-dresden.de), Alexandre Tkatchenko (alexandre.tkatchenko@uni.lu)
[†]Present address: Institute for Materials Science and Max Bergmann Center of Biomaterials, TU Dresden, 01062 Dresden, Germany.

A. Fallani, L. Medrano Sandonas, and A. Tkatchenko,
Nature Commun., *under review* (2023).

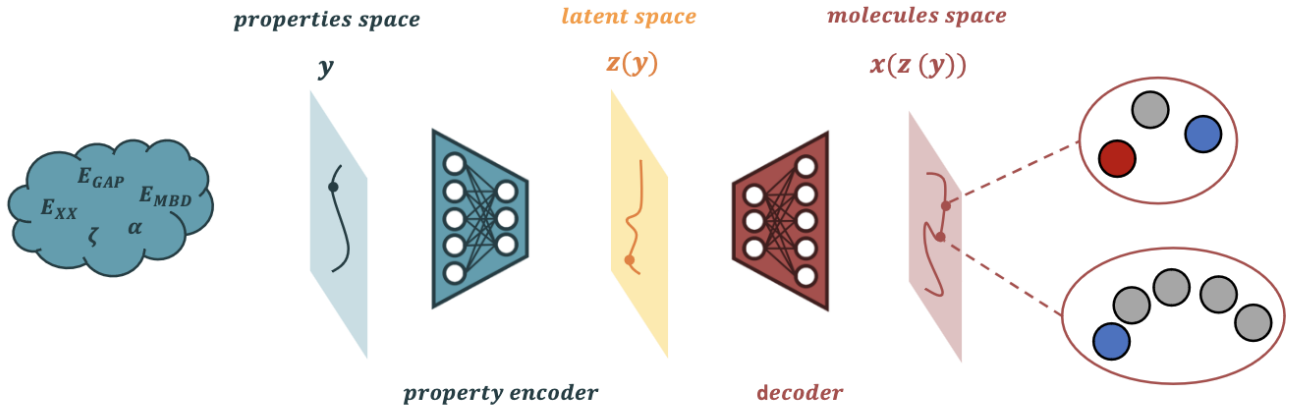
(Baby)-Steps Towards Inverse Molecular Design

(a) TRAINING



$$loss^{joint}_{ELBO} = D_{KL}[q(z|x)||p(z)] - \mathbb{E}_q(z|x)[\ln(p(x|z))] - \mathbb{E}_{q(z|x)}[\ln(h(z|t))]$$

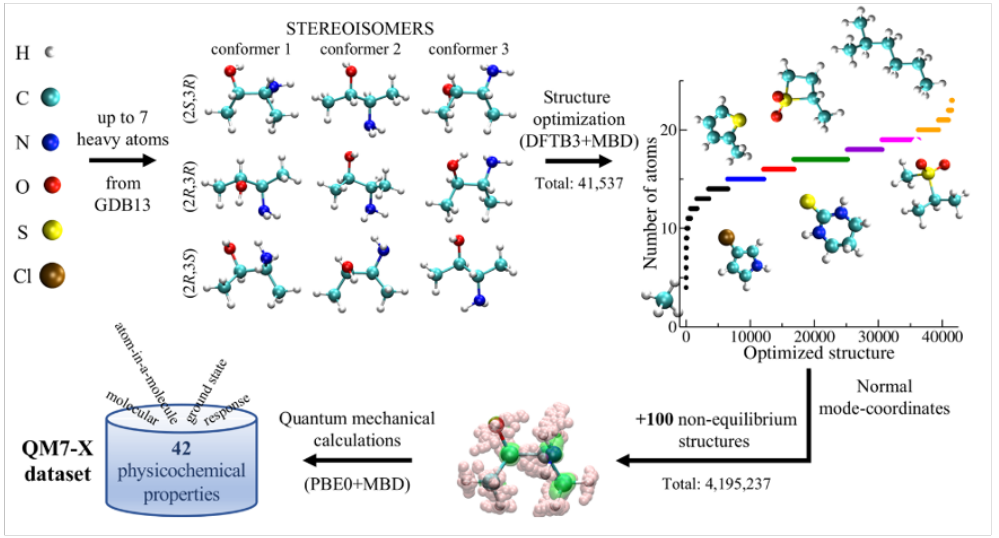
(b) INFERENCE



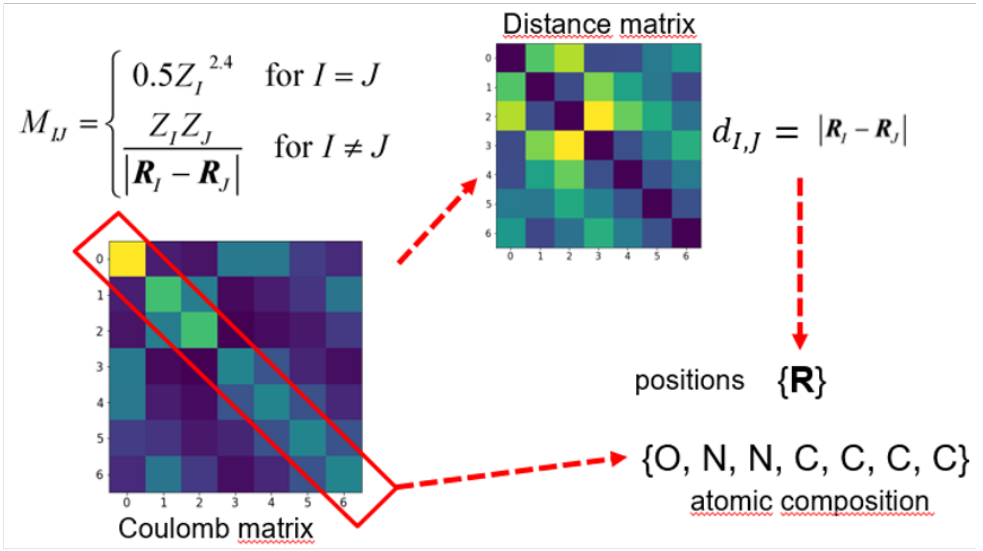
A. Fallani, L. Medrano Sandonas, and A. Tkatchenko,
Nature Commun., *under review* (2023).

(Baby)-Steps Towards Inverse Molecular Design: Dataset and Representation

- We use only equilibrium structures containing H, C, N, O from the QM7-X dataset, and an array of 17 molecular properties



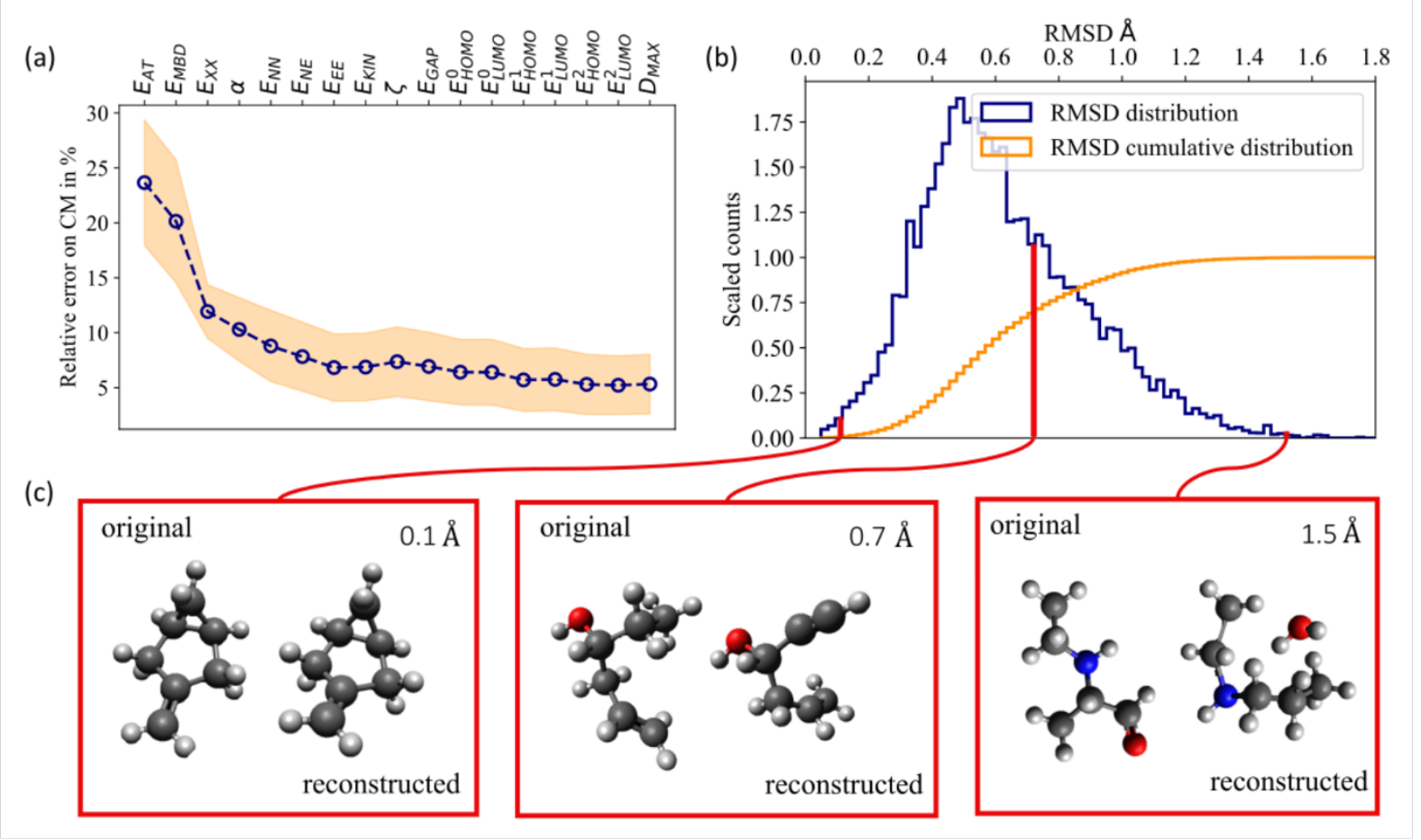
- The representation is a Coulomb matrix without hydrogens, allowing for relative positions and atomic composition retrival



Hoja, J. et al. Sci Data **8,** 43 (2021)

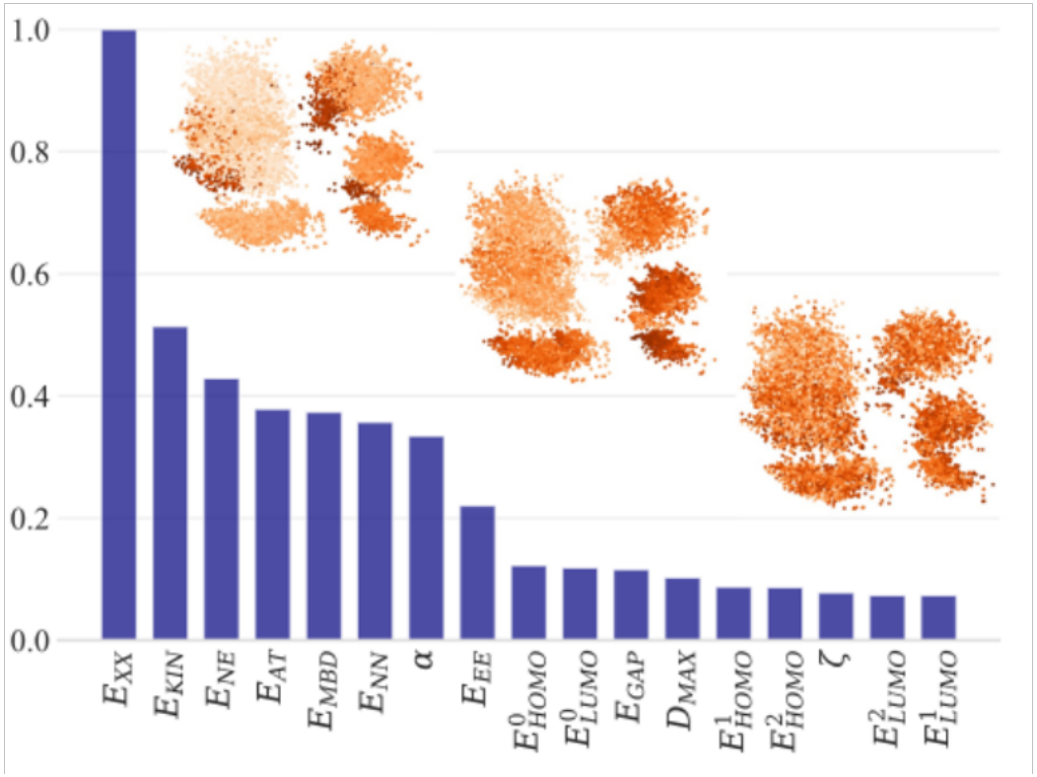
(Baby)-Steps Towards Inverse Molecular Design: Structure Reconstruction from QM Properties

- The Coulomb matrices in the test set are reconstructed within an error close to 5%
- The structures differ by an <RMSD> = 0.6$\dot{A}$ wrt the corresponding molecules in the test set.



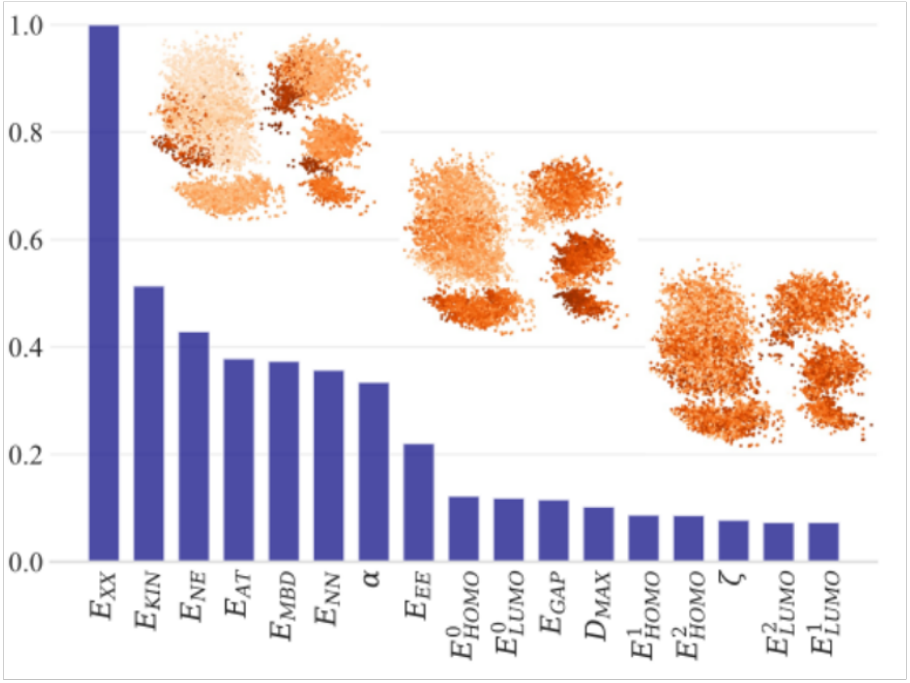
(Baby)-Steps Towards Inverse Molecular Design: Which QM Properties?

- Kinetic and exhange energy seem to be rather important, but why?

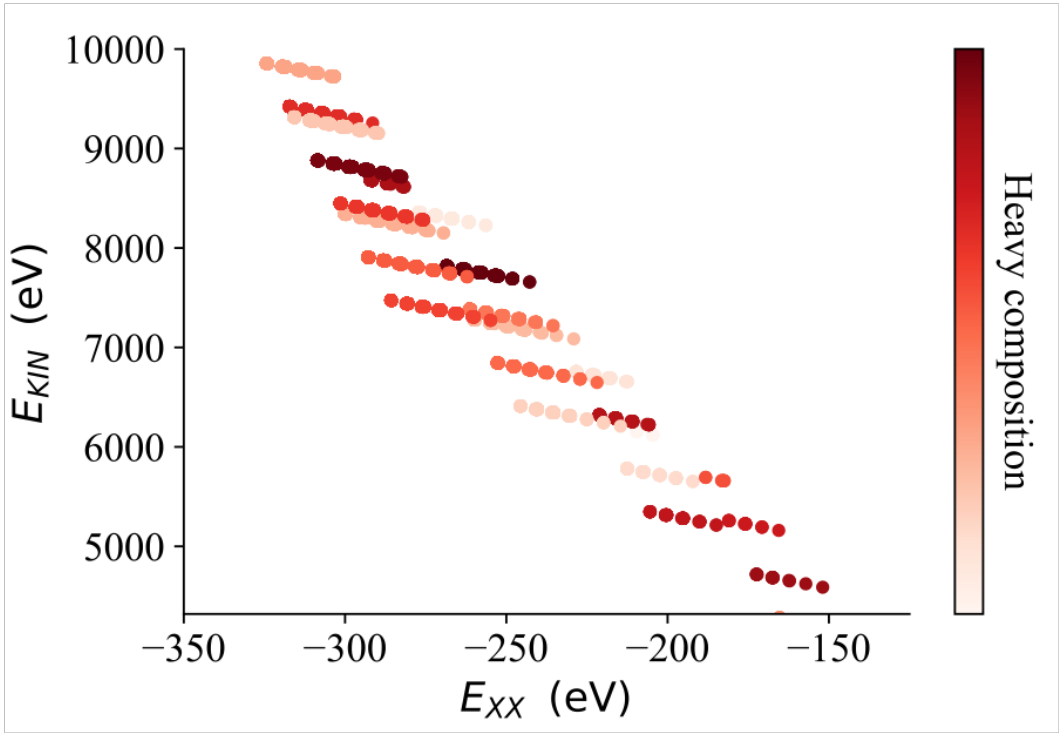


(Baby)-Steps Towards Inverse Molecular Design: Which QM Properties?

- Kinetic and exhange energy seem to be rather important, but why?

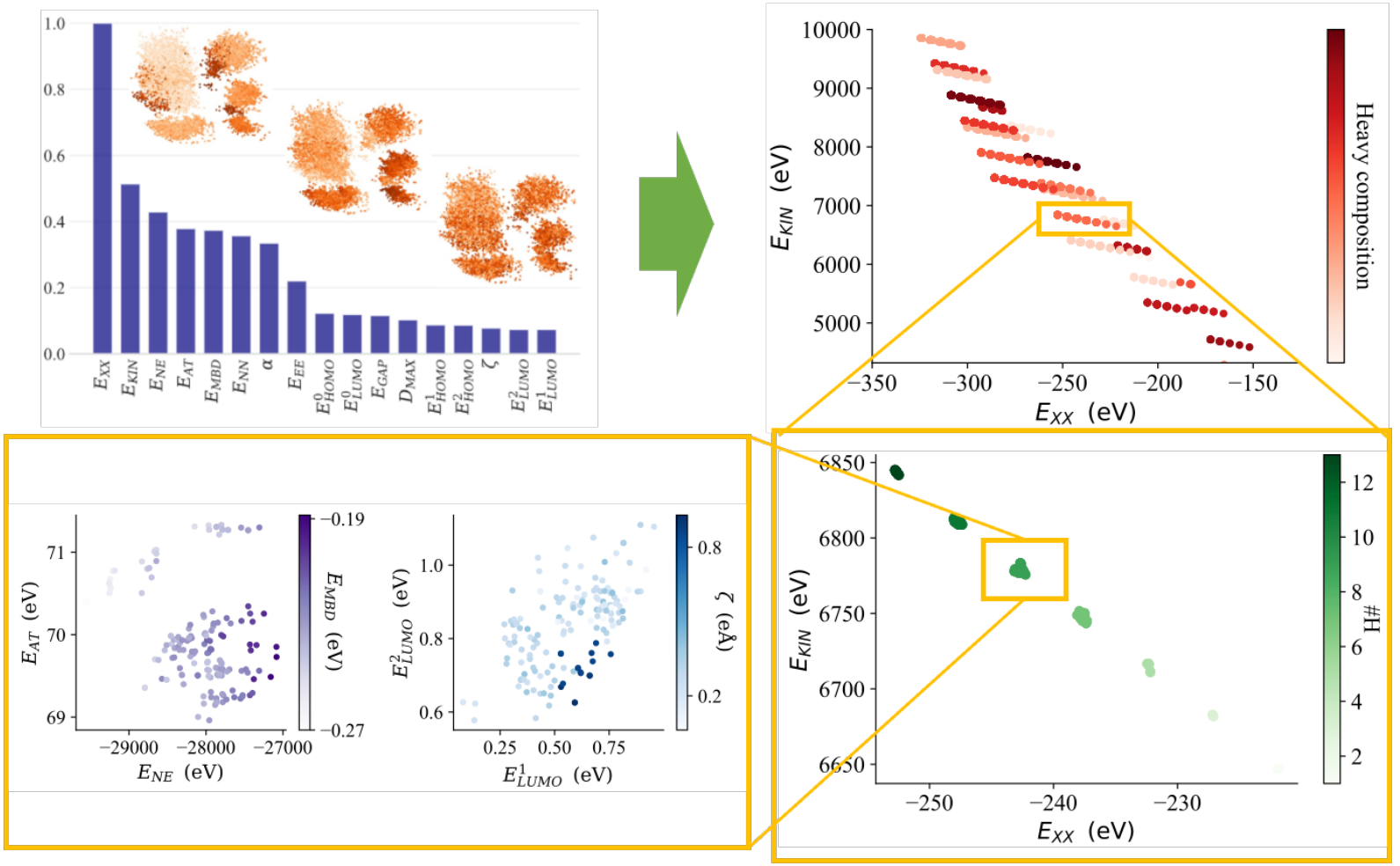


- We find that those properties organize the dataset in *stripes* of molecules with same heavy atom composition



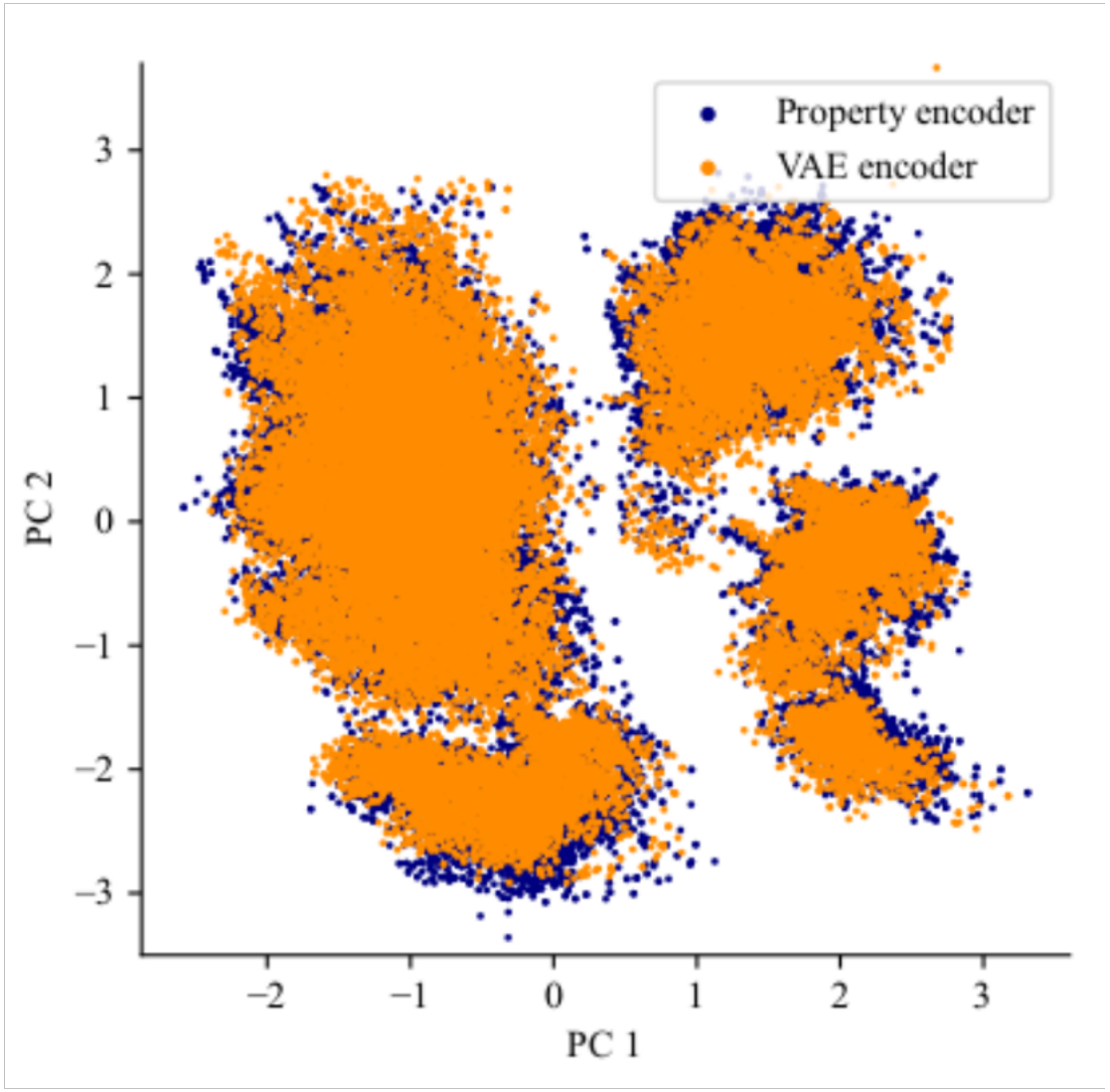
(Baby)-Steps Towards Inverse Molecular Design: Which QM Properties?

- Isolating one of these *stripes* we find it is further clustered based on number of hydrogen atoms
- Each cluster is an isomeric subspace which is better isolated by high gradient map valued properties



(Baby)-Steps Towards Inverse Molecular Design: A Common Latent Space

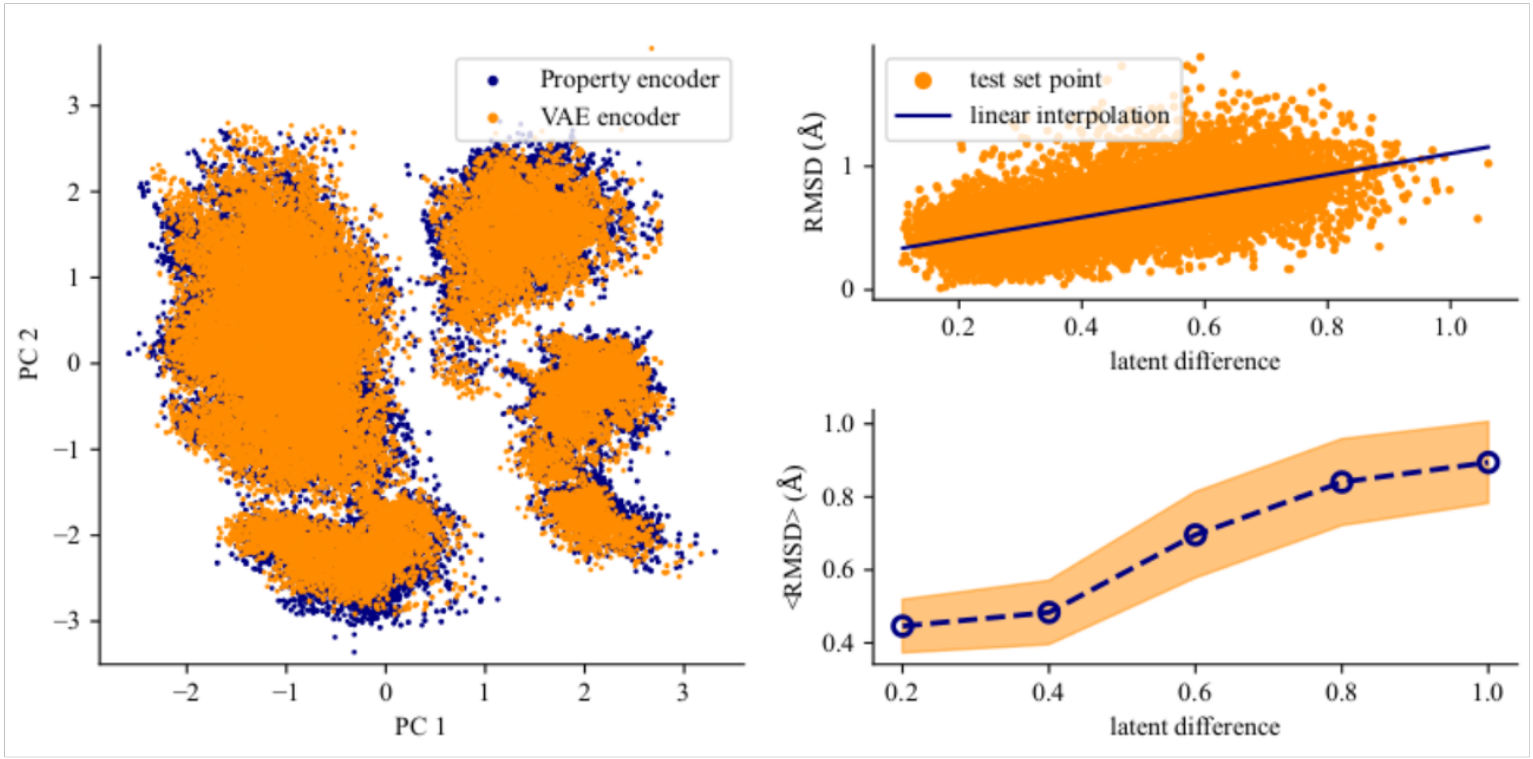
- A PCA of the latent representation of the latent encoding from the VAE (z) and the one from the property encoder ($\tilde{z}$) reveal a very similar structure as expected



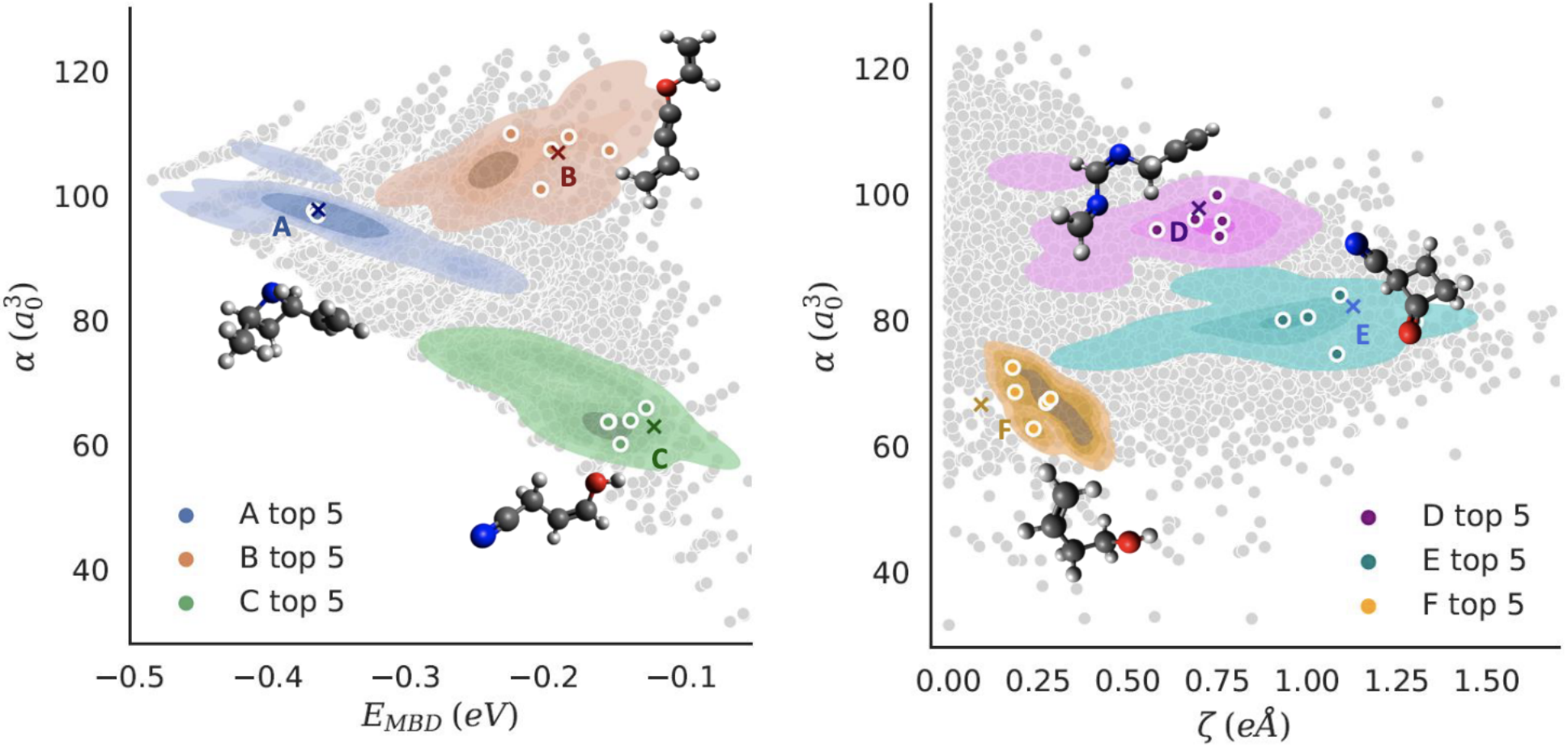
(Baby)-Steps Towards Inverse Molecular Design: A Common Latent Space

- A PCA of the latent representation of the latent encoding from the VAE (Z) and the one from the property encoder ($\tilde{Z}$) reveal a very similar structure as expected

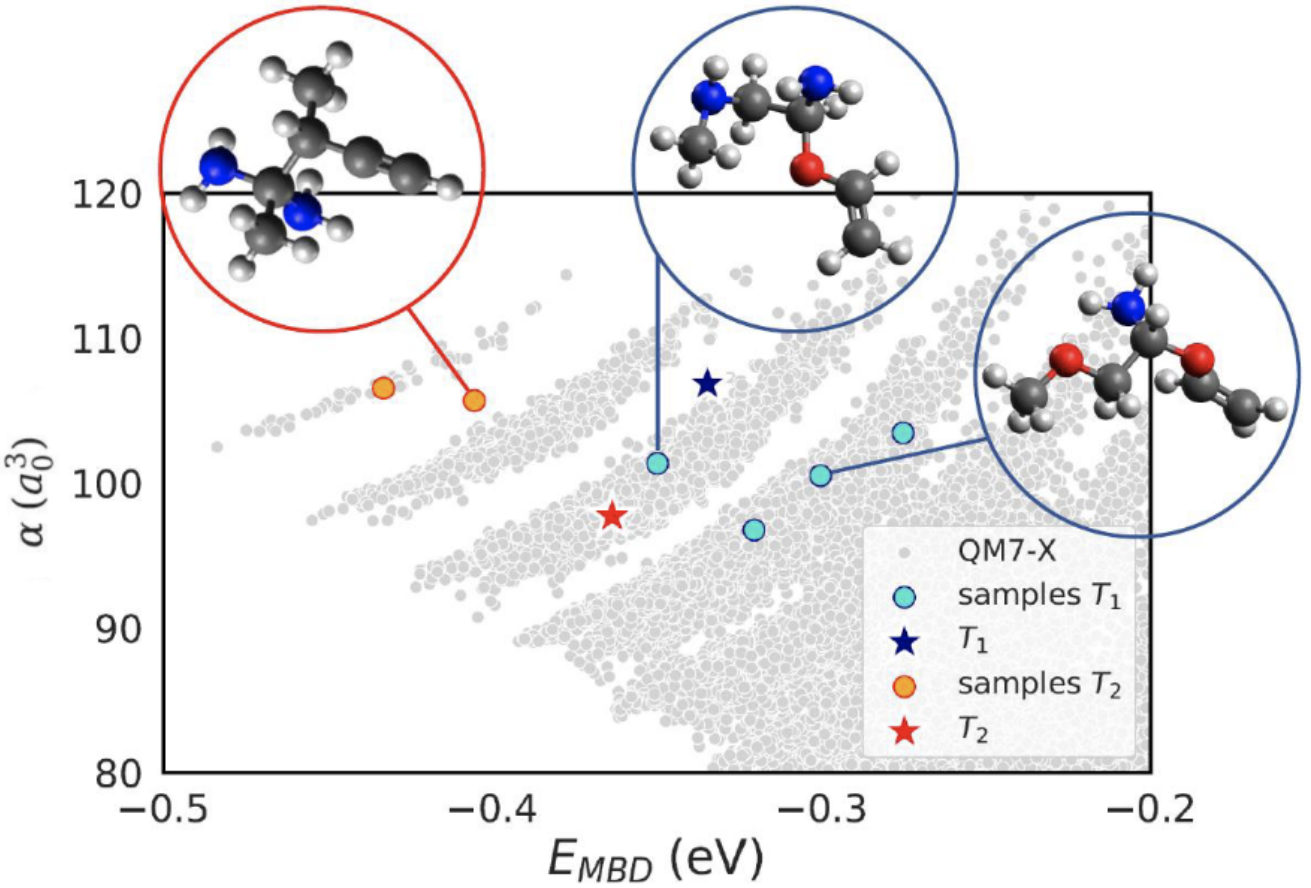
- We find that the *latent difference* $\| Z - \tilde{Z} \|$ correlates with the reconstruction error (RMSD) in the test set



(Baby)-Steps Towards Inverse Molecular Design: Finding Molecules with Given Properties

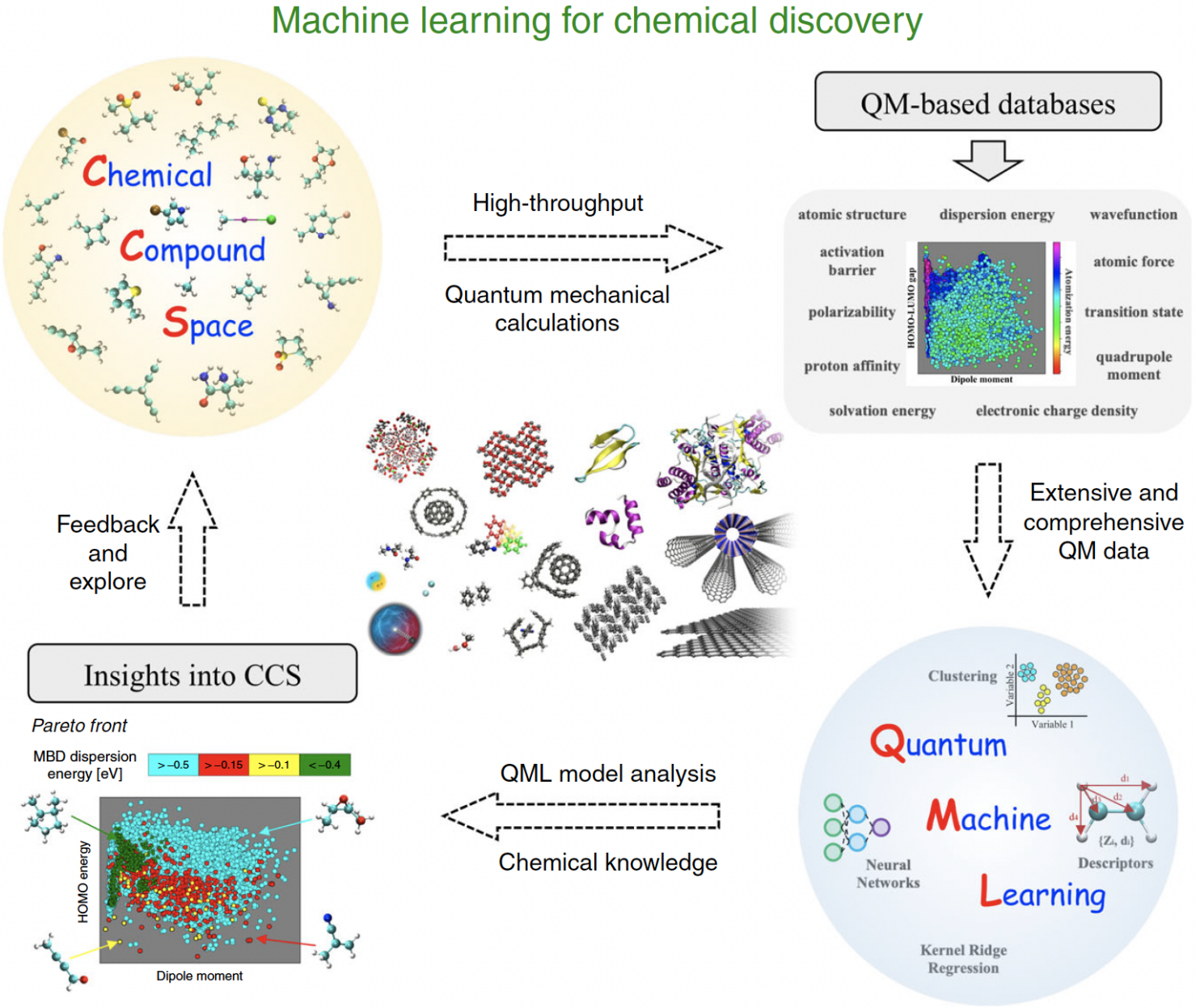


(Baby)-Steps Towards Inverse Molecular Design: ”Extrapolation” to *De Novo* Molecules



A. Fallani, L. Medrano Sandonas, and A. Tkatchenko,
Nature Commun., *under review* (2023).

Combining Direct and Inverse Molecular Design



A. Tkatchenko, ***Nature Commun.*** 11, 4125 (2020).